CSCE 5290: Natural Language Processing

FRAUD DETECTION BY USING TEXT CLASSIFICATION

Project Proposal Description:

1a) Project Title:

We are decided to propose the project of Using the **Fraud Detection by Using the Technique of Text Classification** based on the techniques of Natural Language Processing like Text classification and Named entity Recognition.

1b) **Team Members:**

   Group-07

1.Venkata Manisahith

2.Lokesh Naidu Bavigadda

3.Srikanth Karni

4.Pasumarthi Raghu Ram

GitHub Link: Link

Goals and Objectives:

• Motivation:

➢ The motivation for this project is to improve the efficiency and accuracy of existing traditional fraud detection techniques.

Generally, many of us observe negative reviews for the products and movies, spam Mails related to many companies. Fraudulent activities can cause significant financial due to the loss of reputation due to these activities.

➢ The traditional fraud detection techniques may not be able to keep up with evolving techniques as the fraudsters may be able to develop some alternative software to overcome the existing algorithms.

➢ So, by using both Natural Language Processing and Machine learning techniques the organizations can be able to classify the text data like customer reviews, emails and transaction descriptions whether they are fake or original automatically.

➢ Also, this process is very fast so that the organizations can act very fast so that they can stop huge reach among their customers. This can help them to maintain customer trust.

• **Significance:**

We can say   significance of this project mainly by using four parameters   like

1.Detecting the frauds in real time to control the damage as much as possible.

2.Financial loss can be controlled by this project as it controls the damage so that the specific company can damage control processes like refunding to the customers, decreasing the prices to increase customers for the product in less number.

3.It can detect the frauds like spam mails and negative reviews by the same members very accurately.

4.It is the important parameter called Scalability. As the companies expand their operations, traditional fraud detection methods may struggle to cop up with the increased volume of data. This text

classification can be able to process the increasing volume of data by setting some parameter while designing the algorithm.

- **Objectives:**

  ➢ By considering the   common problems we decided to make the project as its objectives as

  ➢ To save time and resources: Traditional fraud detection techniques can be time consuming and use many resources which can be expensive process with addition they loss the customer which means brand damage. Here our project automates the process very fast in real time.

  ➢ To reduce Fraudulent activities in number: Due its accuracy and fastness it can be able to reduce the activities by fraudsters by using this algorithm.

- **Features:**

  Data Preprocessing:

  ➢ A key feature of any text classification technique is the ability to preprocess and clean the text data before analysis.

  ➢ This include the tasks like removing stop words, stemming, lemmatization, and removing punctuations in the sentence which can be easy for tokenization.

  Feature Extraction:

  ➢ It is ability to extract relevant features from the text data.

  ➢ Tis can include techniques like bag of words, word embeddings to convert the text data into numerical vectors that can be used as input for machine learning models.

  Integration with other systems:

  ➢ To enable seamless integration with other fraud detection systems and prevention systems, organizations we want to include features

that allow for easy integration. This can include API's, web hooks etc.,

References:

Here are some of the references that can be useful for a text classification project for fraud detection:

➢ Zhang, X., Li, Q., Huang, X., & Wu, X. (2018). A deep learning approach for financial fraud detection based on textual data. Expert Systems with Applications, 107, 12-21.
➢ Abbasi, S. A., & Chen, H. (2008). Cybercrime classification: A motivational model. Communications of the ACM, 51(3), 88-93
➢ Wang, Z., Sun, Y., & Gao, X. (2020). A novel credit card fraud detection model based on word embedding and deep learning. Expert Systems with Applications, 156, 113397.
➢ Kumar, N., Agarwal, A., & Choudhary, A. (2020). Fraud Detection in E-commerce: A Comprehensive Review. Journal of Ambient Intelligence and Humanized Computing, 11(10), 4235-4261